PROJECT DOCUMENTATION

Introduction

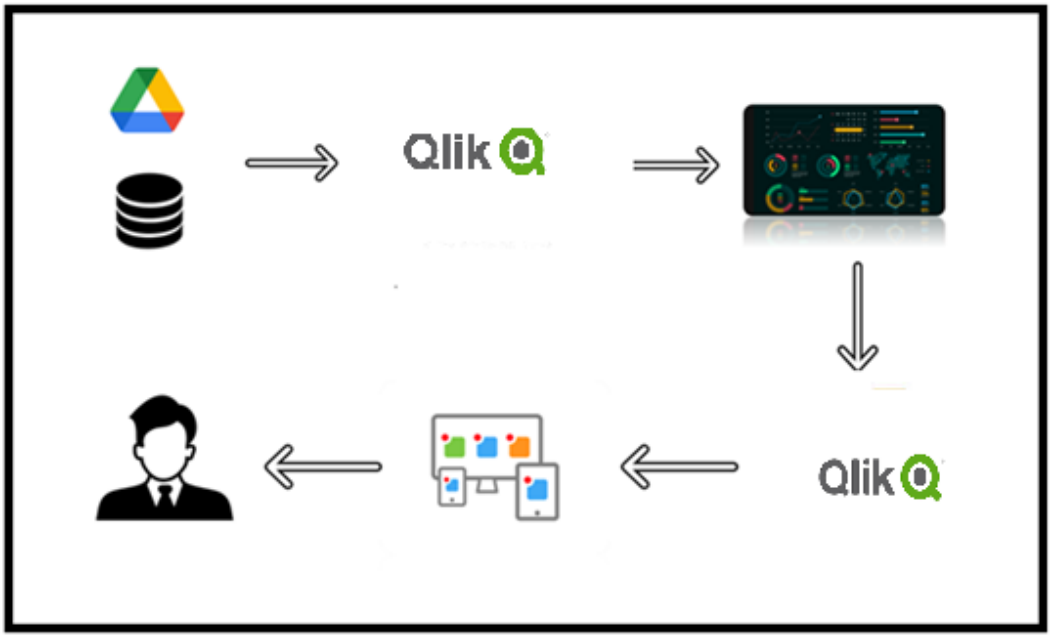
1.1 Overview

The project aims to conduct a comprehensive analysis of peer-to-peer lending data from LendingClub to gain valuable insights into loan performance, borrower behavior, and associated risk factors. By leveraging a robust dataset, the analysis will explore various aspects of the lending process, including the factors that influence loan defaults, the characteristics of borrowers who successfully repay their loans, and the overall trends in borrower demographics and loan types. The goal is to identify key patterns and correlations that can inform better decision-making for investors and policymakers within the peer-to-peer lending ecosystem. Through statistical analysis, machine learning techniques, and data visualization, the project seeks to enhance the understanding of risk management and optimize the lending strategies for improved financial outcomes.

1.2 Purpose

The purpose of this project is to perform an in-depth analysis of LendingClub's peer-to-peer lending data to extract meaningful insights into loan performance, borrower behavior, and risk factors. The project aims to identify the determinants of successful loan repayments and defaults, understand borrower demographics and their impact on lending outcomes, and uncover risk factors that influence loan performance. By utilizing advanced analytical methods, the project seeks to provide actionable recommendations for investors to optimize their lending strategies and for policymakers to enhance regulatory frameworks. Ultimately, the project aims to contribute to the efficiency and reliability of the peer-to-peer lending market, fostering better investment decisions and more effective risk management.

1.3 Technical Architecture



2 Define Problem

2.1 Specify the business problem

The specific business problem revolves around the inadequacy of the current lending strategy, which is not sufficiently informed by comprehensive insights derived from LendingClub loan data. The institution struggles to assess borrower behavior and market dynamics effectively, resulting in challenges such as inaccurate risk identification, difficulties in predicting loan default rates, and the inability to dynamically adjust lending criteria to respond to evolving market conditions. Furthermore, this lack of detailed analysis leads to suboptimal portfolio performance, reduced investor confidence, and potential financial losses. Without a data-driven approach, the institution is also unable to leverage emerging trends and innovate its lending practices to maintain competitiveness in the peer-to-peer lending market. Addressing these issues through a thorough analysis of LendingClub data can pave the way for more precise risk management, enhanced predictive capabilities, and a more adaptable and resilient lending strategy. By integrating comprehensive insights into their operations, the institution can improve loan approval processes, tailor financial products to better meet borrower needs, and ultimately achieve a more robust and sustainable growth trajectory in the lending landscape.

2.2 Business Requirements

The business requirements involve the establishment of a robust data analytics framework that can extract meaningful insights from LendingClub issued loans data. This framework should enable the financial institution to gain a deep understanding of borrower behavior, identify high-risk segments, predict default rates accurately, and provide the necessary foundation for real-time adjustments to lending criteria. Additionally, the solution should be scalable, adaptable, and capable of integrating with existing systems to ensure seamless

implementation. The analytics framework should also include advanced visualization tools to facilitate intuitive data interpretation and reporting, empowering decision-makers with actionable intelligence. Furthermore, it should support continuous learning and improvement by incorporating machine learning algorithms that refine predictions and risk assessments over time. The framework must also ensure data security and compliance with relevant regulations, safeguarding sensitive borrower information. By meeting these requirements, the financial institution can enhance its strategic decision-making processes, improve loan performance, and achieve a competitive edge in the peer-to-peer lending market.

2.3 Literature Survey

Peer-to-peer (P2P) lending platforms like LendingClub have transformed finance by directly connecting borrowers and investors, bypassing traditional intermediaries. Big data and advanced analytics have further enhanced this market, enabling detailed analyses of loan performance, borrower behavior, and associated risks.

Studies have developed models to predict loan defaults using machine learning. Serrano-Cinca et al. (2015) used logistic regression and decision trees, finding key borrower characteristics like credit score and debt-to-income ratio impact default probabilities. Xia et al. (2017) proposed a dynamic risk assessment framework incorporating borrower credit information and macroeconomic factors. Malekipirbazari and Aksakalli (2015) found ensemble methods, particularly random forests, to be superior in predictive performance.

Research on borrower behavior identified determinants of successful loan acquisition. Herzenstein et al. (2008) found that both qualitative and quantitative borrower attributes, such as narrative quality and financial history, are crucial. Lin et al. (2013) showed that borrowers with social ties to investors had lower default rates. Pope and Sydnor (2011) demonstrated distinct borrowing behaviors and risk profiles across different borrower segments.

Data analytics is crucial in P2P lending. Kshetri (2017) highlighted its role in enhancing platform capabilities with real-time analytics. Yildirim et al. (2018) reviewed machine learning applications, emphasizing supervised learning for predicting defaults and unsupervised learning for segmentation. Chen et al. (2019) proposed visual analytics dashboards for real-time monitoring.

Challenges like data quality and integration remain critical. Zhang et al. (2016) emphasized the need for accurate datasets. Regulatory compliance is another challenge, as noted by Wadhwa et al. (2019), requiring integration into analytics frameworks. Ethical considerations in data usage, as discussed by Binns (2018), are essential for fairness and transparency.

3 Data Collection

3.1 Collect the Dataset

\

To conduct a comprehensive analysis of LendingClub's peer-to-peer lending data, the first step is to collect the relevant dataset. This involves accessing LendingClub's publicly available datasets, which provide detailed information on loan applications, approved loans, borrower demographics, loan statuses, and financial metrics. Visit LendingClub's website or data repository, select and download the necessary data files in a usable format like CSV or Excel. Review the accompanying documentation to

understand data structure and definitions. Verify the integrity and completeness of the datasets, ensuring no missing or corrupted entries. Finally, organize and securely store the data for subsequent analysis.

3.2 Connect Data with Qlik Sense

The dataset has been successfully connected with Qlik Sense, enabling seamless integration of data analytics and visualization capabilities. After preparing the dataset in compatible formats like CSV or Excel, the Data Manager or Data Load Editor in Qlik Sense was utilized to establish the connection. Following this, the data was loaded into Qlik Sense, and data models were created for analysis. Using the intuitive interface of Qlik Sense, interactive visualizations and dashboards were crafted to explore insights within the dataset. The findings have been shared with colleagues or stakeholders through published apps on Qlik Sense Server or Qlik Cloud, facilitating collaborative decision-making based on data-driven insights.

4 Data Preparation

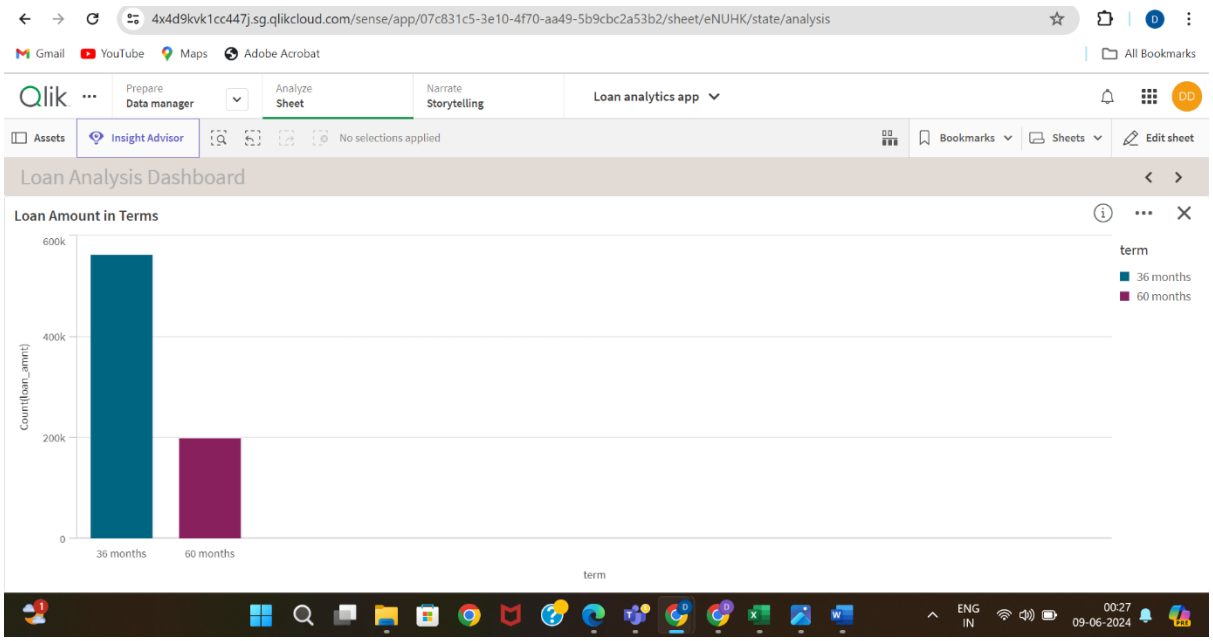
4.1 Prepare the data for Visualisation

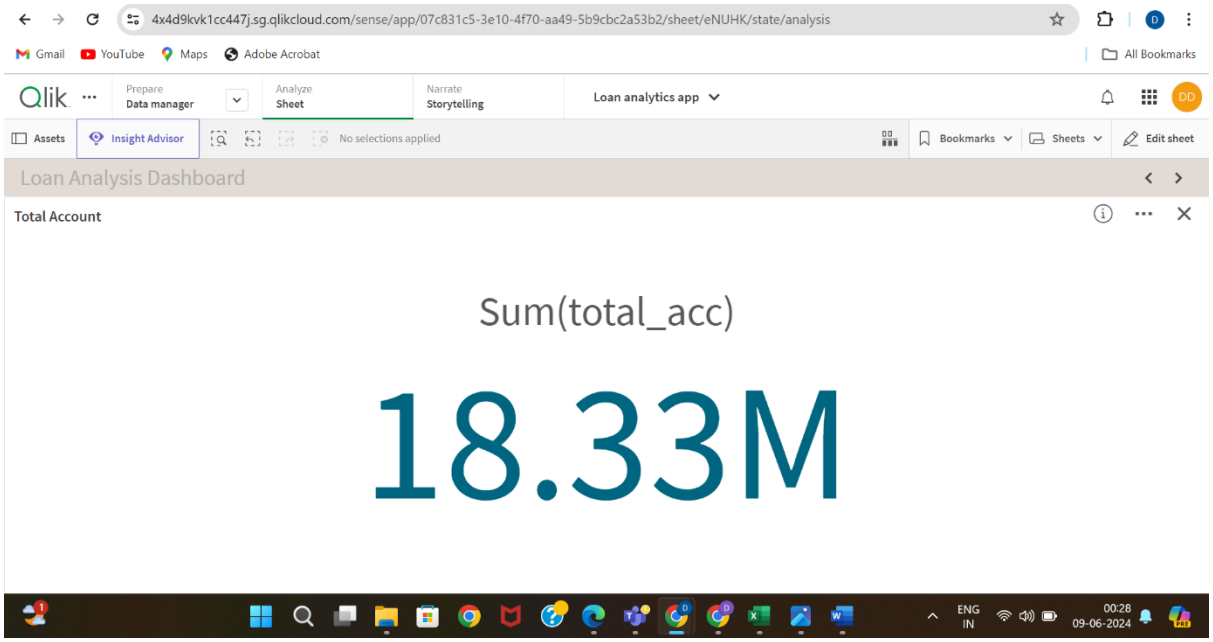
The data has undergone thorough processing and preparation to ensure it is ready for analysis. Initially, the dataset was formatted into compatible types such as CSV or Excel. It was then cleansed to remove any inaccuracies, inconsistencies, and missing values. Transformations were applied to standardize and normalize the data, ensuring it meets the necessary quality standards. Additionally, relevant features were selected and data models were created to facilitate effective analysis. This comprehensive data processing and preparation ensure that the data is accurate, consistent, and ready for insightful analysis.

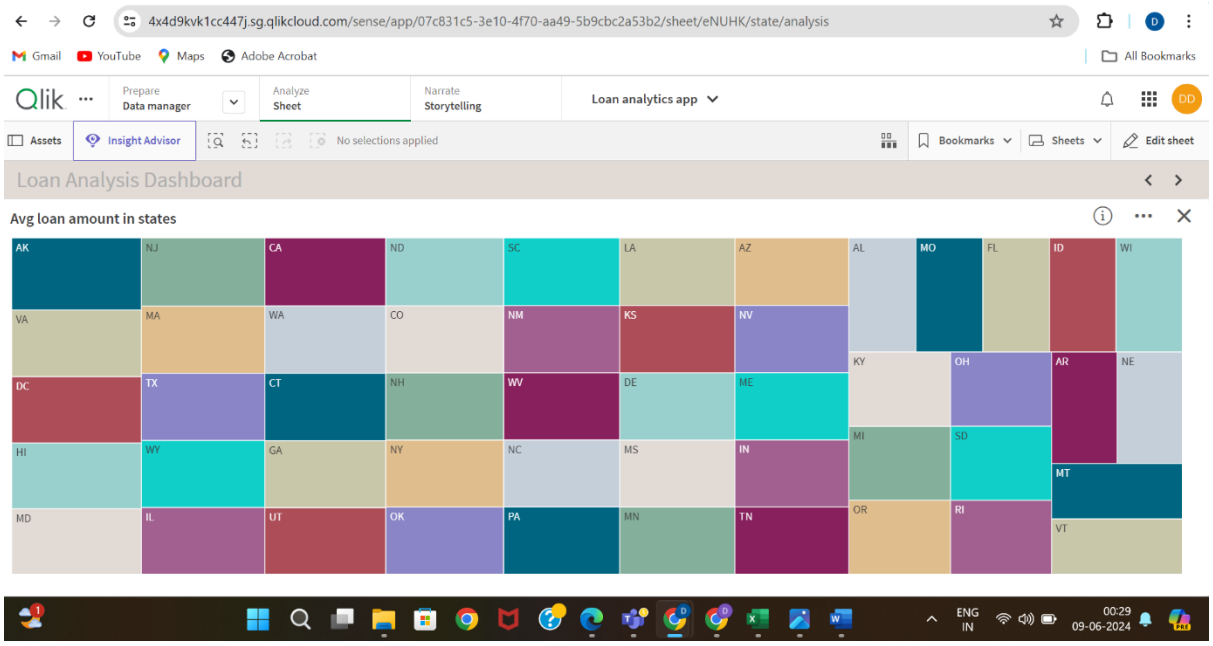
5 Data Visualizations

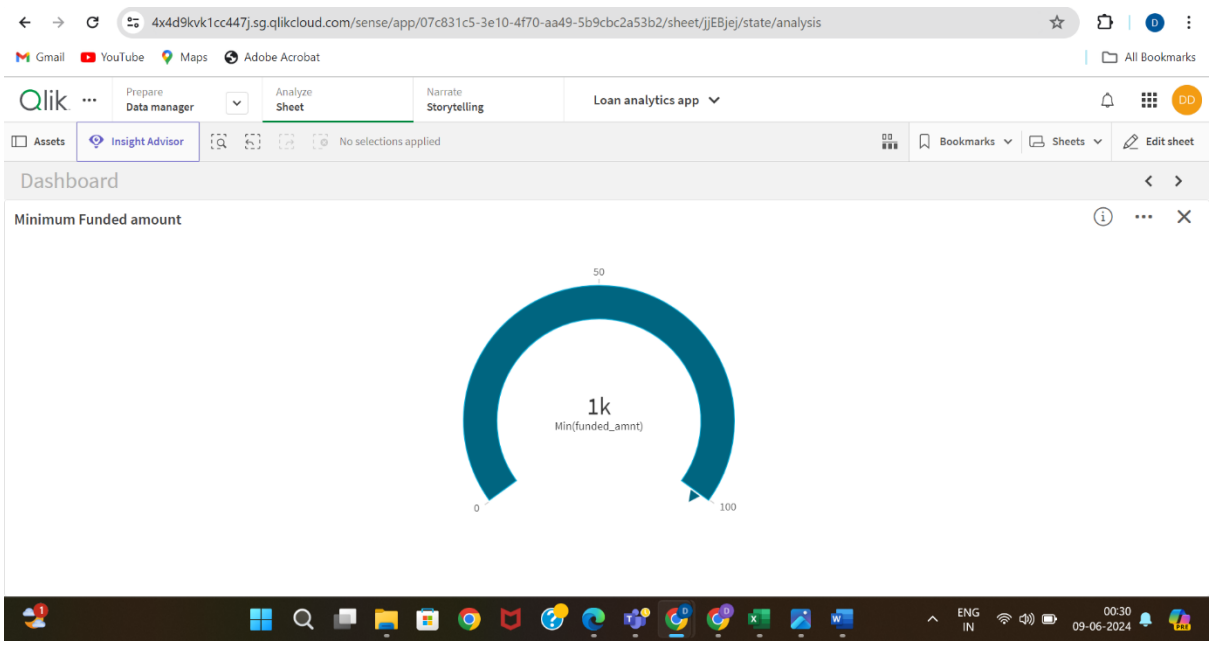
5.1 Visualizations

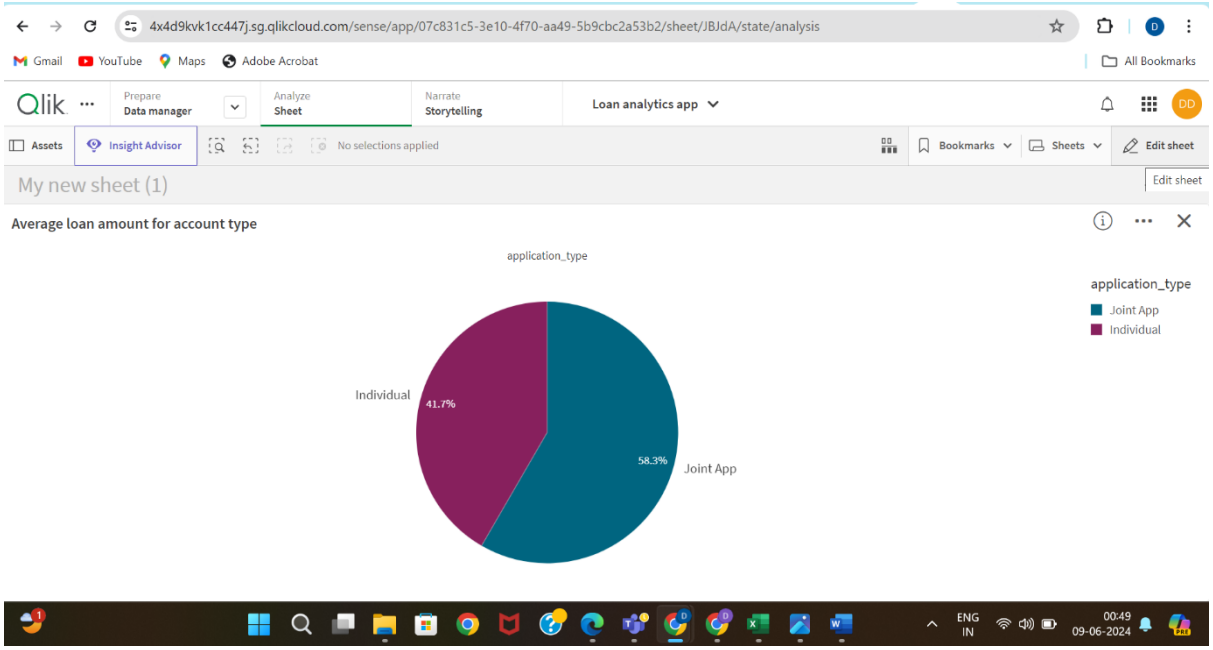
Data visualizations in data analysis are graphical representations of data designed to make complex information easier to understand. They include charts, graphs, maps, and dashboards that display data patterns, trends, and insights visually. Common types of visualizations are bar charts, line graphs, pie charts, scatter plots, and heat maps. These tools help analysts and decision-makers quickly grasp large datasets, identify relationships, and derive actionable insights.

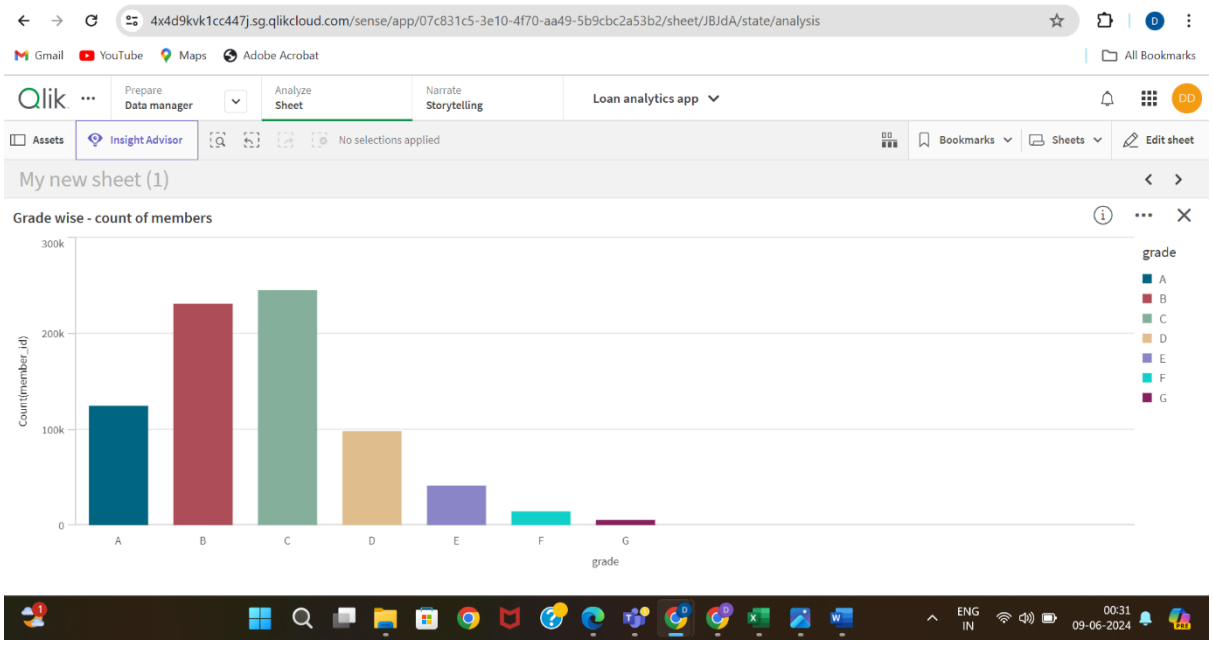

4x4d9kvk1cc447j.sg.qlikcloud.com/sense/app/07c831c5-3e10-4f70-aa49-5b9cbc2a53b2/sheet/eNUHK/state/analysis
Gmail
YouTube
Maps
Adobe Acrobat
All Bookmarks
Qlik
Prepare
Data manager
Analyze
Sheet
Narrate
Storytelling
Loan analytics app
DD
Assets
Insight Advisor
No selections applied
Bookmarks
Sheets
Edit sheet
Loan Analysis Dashboard
Loan Amount in Terms
600k
400k
200k
0
Count(loan_amnt)
36 months
60 months
term
term
36 months
60 months
ENG
IN
00:27
09-06-2024

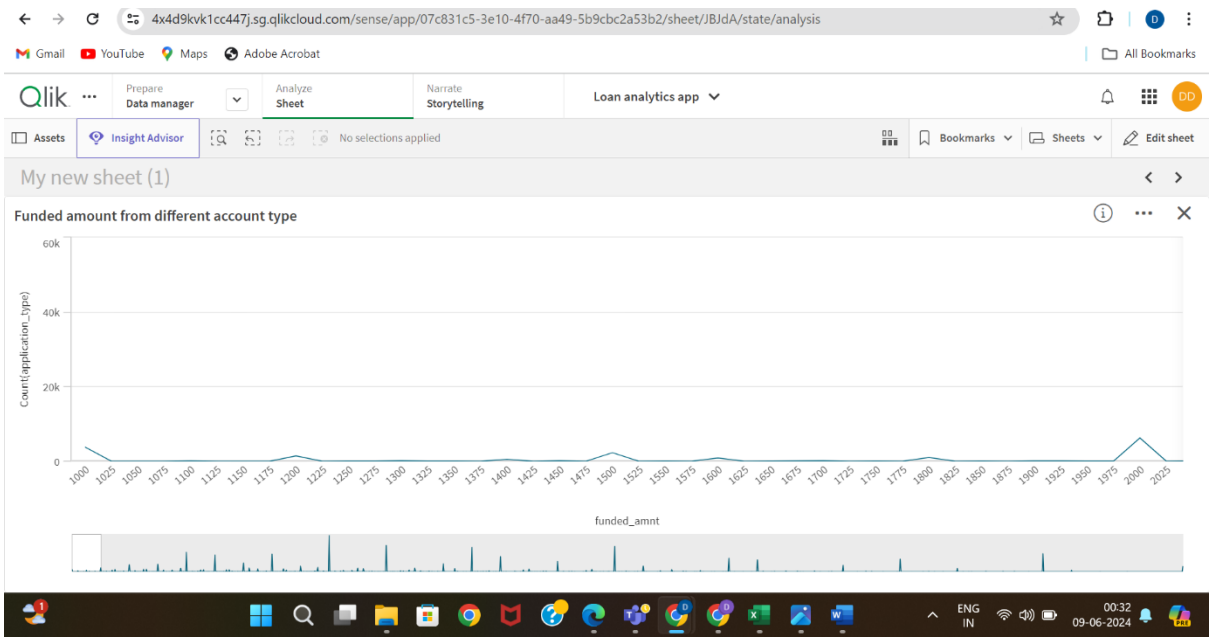

4x4d9kvk1cc447j.sg.qlikcloud.com/sense/app/07c831c5-3e10-4f70-aa49-5b9cbc2a53b2/sheet/eNUHK/state/analysis
Gmail
YouTube
Maps
Adobe Acrobat
All Bookmarks
Qlik
Prepare
Data manager
Analyze
Sheet
Narrate
Storytelling
Loan analytics app
DD
Assets
Insight Advisor
No selections applied
Bookmarks
Sheets
Edit sheet
Loan Analysis Dashboard
Total Account
Sum(total_acc)
18.33M
ENG
IN
00:28
09-06-2024

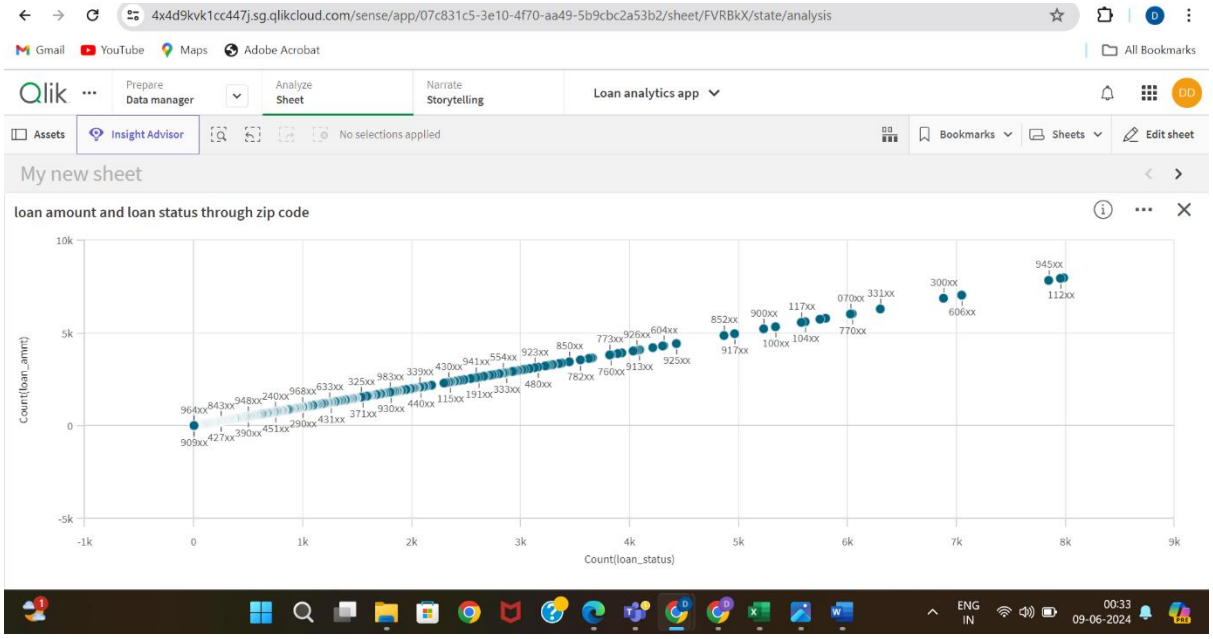

4x4d9kvk1cc447j.sg.qlikcloud.com/sense/app/07c831c5-3e10-4f70-aa49-5b9cbc2a53b2/sheet/eNUHK/state/analysis
Gmail
YouTube
Maps
Adobe Acrobat
All Bookmarks
Prepare
Data manager
Analyze
Sheet
Narrate
Storytelling
Loan analytics app
Assets
Insight Advisor
No selections applied
Bookmarks
Sheets
Edit sheet
Loan Analysis Dashboard
Avg loan amount in states
AK
NJ
CA
ND
SC
LA
AZ
AL
MO
FL
ID
WI
VA
MA
WA
CO
NM
KS
NV
KY
OH
AR
NE
DC
TX
CT
NH
WV
DE
ME
MI
SD
HI
WY
GA
NY
NC
MS
IN
MT
MD
IL
UT
OK
PA
MN
TN
OR
RI
VT
ENG
IN
00:29
09-06-2024

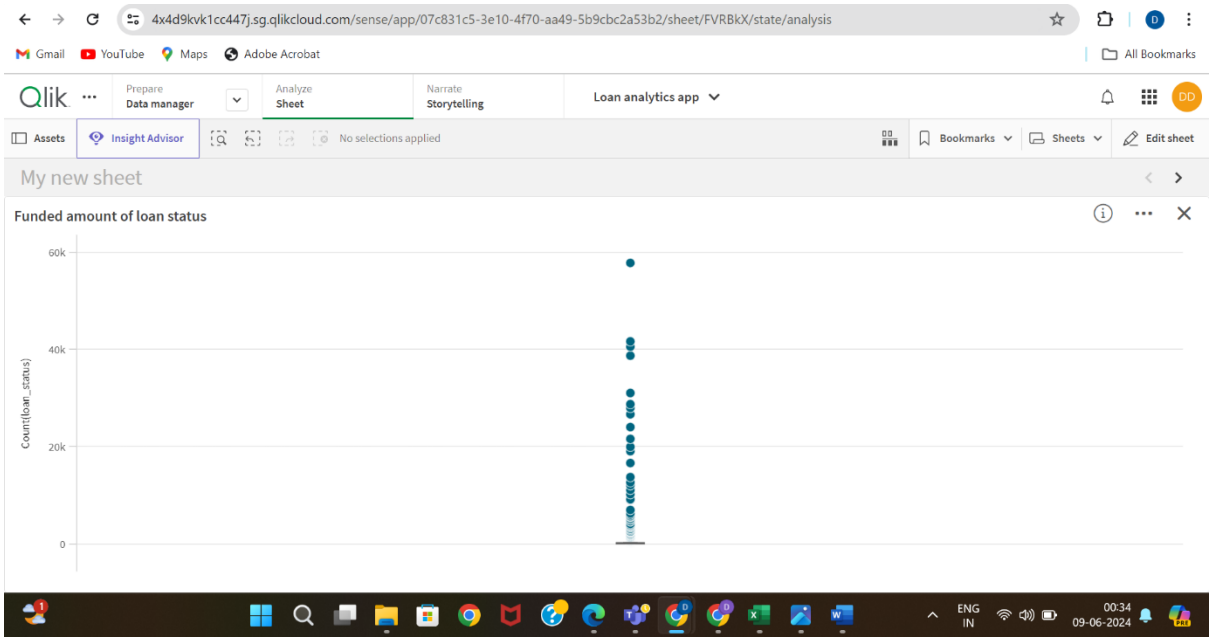

4x4d9kvk1cc447j.sg.qlikcloud.com/sense/app/07c831c5-3e10-4f70-aa49-5b9cbc2a53b2/sheet/jjEBjej/state/analysis
Gmail
YouTube
Maps
Adobe Acrobat
All Bookmarks
Prepare
Data manager
Analyze
Sheet
Narrate
Storytelling
Loan analytics app
Assets
Insight Advisor
No selections applied
Bookmarks
Sheets
Edit sheet
Dashboard
Minimum Funded amount
50
1k
Min(funded_amnt)
0
100
ENG
IN
00:30
09-06-2024


4x4d9kvk1cc447j.sg.qlikcloud.com/sense/app/07c831c5-3e10-4f70-aa49-5b9cbc2a53b2/sheet/JBJdA/state/analysis
Gmail
YouTube
Maps
Adobe Acrobat
All Bookmarks
Qlik
Prepare
Data manager
Analyze
Sheet
Narrate
Storytelling
Loan analytics app
Assets
Insight Advisor
No selections applied
Bookmarks
Sheets
Edit sheet
My new sheet (1)
Edit sheet
Average loan amount for account type
application_type
application_type
Joint App
Individual
Individual
41.7%
58.3%
Joint App
ENG
IN
00:49
09-06-2024


4x4d9kvk1cc447j.sg.qlikcloud.com/sense/app/07c831c5-3e10-4f70-aa49-5b9cbc2a53b2/sheet/JBJdA/state/analysis
Gmail
YouTube
Maps
Adobe Acrobat
All Bookmarks
Qlik
Prepare
Data manager
Analyze
Sheet
Narrate
Storytelling
Loan analytics app
Assets
Insight Advisor
No selections applied
Bookmarks
Sheets
Edit sheet
My new sheet (1)
Grade wise - count of members
300k
200k
100k
0
Count(member_id)
A
B
C
D
E
F
G
grade
grade
A
B
C
D
E
F
G
ENG
IN
00:31
09-06-2024


4x4d9kvk1cc447j.sg.qlikcloud.com/sense/app/07c831c5-3e10-4f70-aa49-5b9cbc2a53b2/sheet/JBJdA/state/analysis
Gmail
YouTube
Maps
Adobe Acrobat
All Bookmarks
Prepare
Data manager
Analyze
Sheet
Narrate
Storytelling
Loan analytics app
Assets
Insight Advisor
No selections applied
Bookmarks
Sheets
Edit sheet
My new sheet (1)
Funded amount from different account type
Count(application_type)
60k
40k
20k
0
funded_amnt
ENG
IN
00:32
09-06-2024

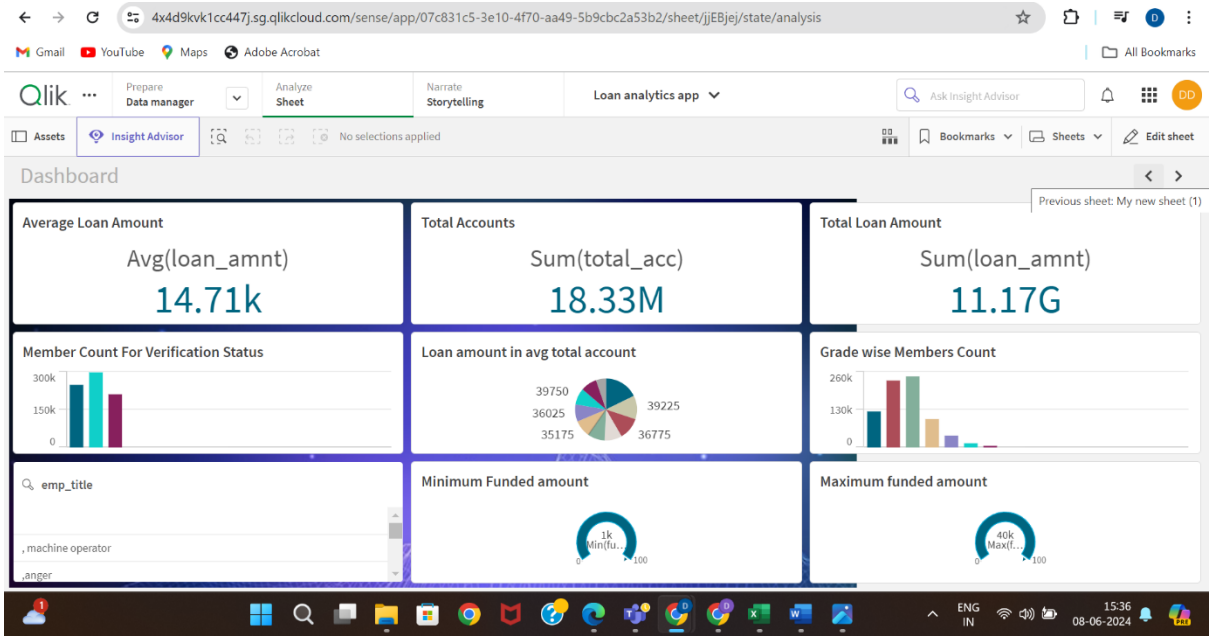

4x4d9kvk1cc447j.sg.qlikcloud.com/sense/app/07c831c5-3e10-4f70-aa49-5b9cbc2a53b2/sheet/FVRBkX/state/analysis
Gmail
YouTube
Maps
Adobe Acrobat
All Bookmarks
Prepare
Data manager
Analyze
Sheet
Narrate
Storytelling
Loan analytics app
Assets
Insight Advisor
No selections applied
Bookmarks
Sheets
Edit sheet
My new sheet
loan amount and loan status through zip code
Count(loan_amnt)
10k
5k
0
-5k
-1k
0
1k
2k
3k
4k
5k
6k
7k
8k
9k
Count(loan_status)
ENG
IN
00:33
09-06-2024



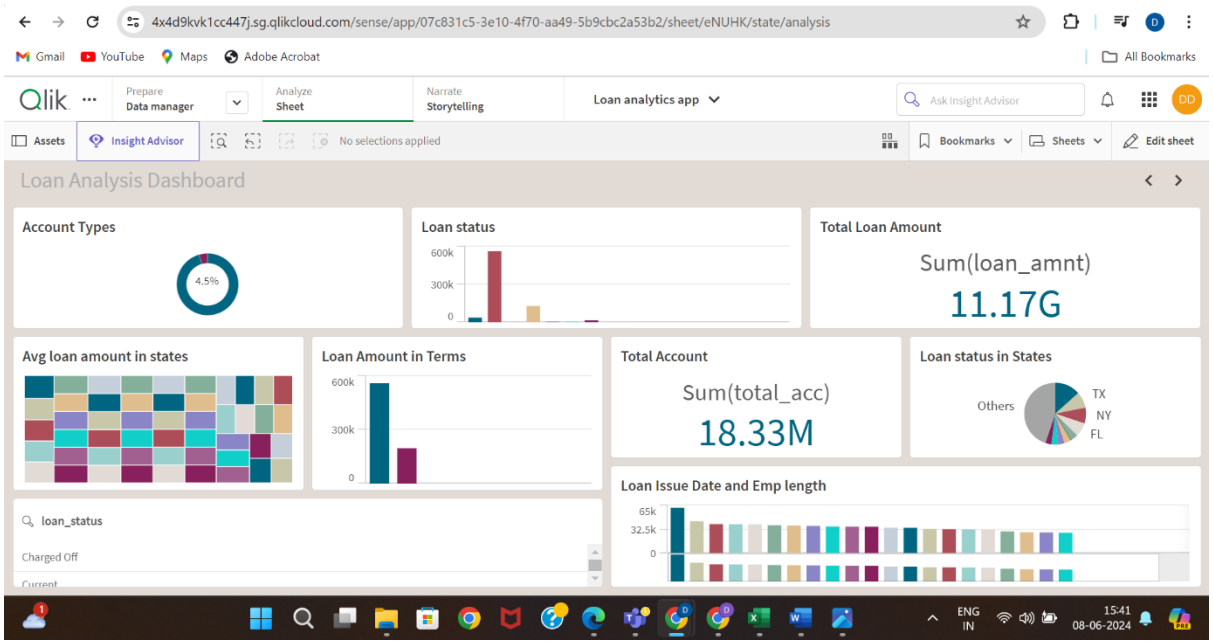
6 Dashboard

A dashboard in data analysis is an interactive, visual display of key metrics and data points, organized on a single screen for easy monitoring and analysis. It consolidates and presents information using charts, graphs, and tables, allowing users to quickly understand performance, trends, and insights. Dashboards are designed to provide real-time data updates and can be customized to show various aspects of the data, making them a powerful tool for informed decision-making and performance tracking.

Dashboard 1



Dashboard 2

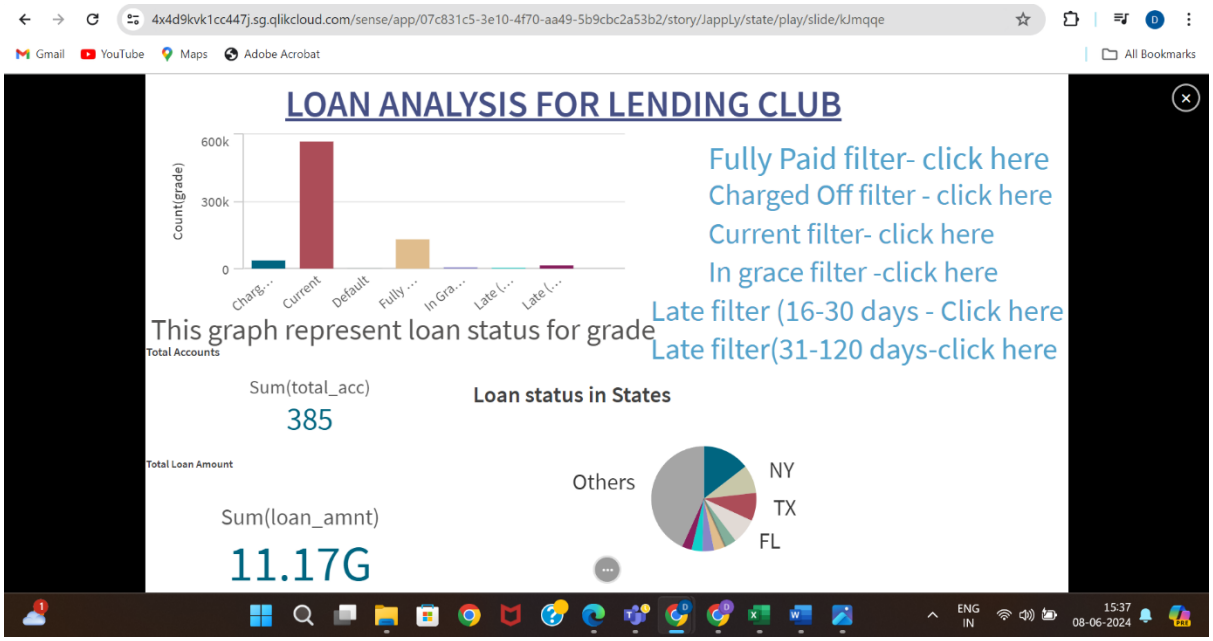


7 Story

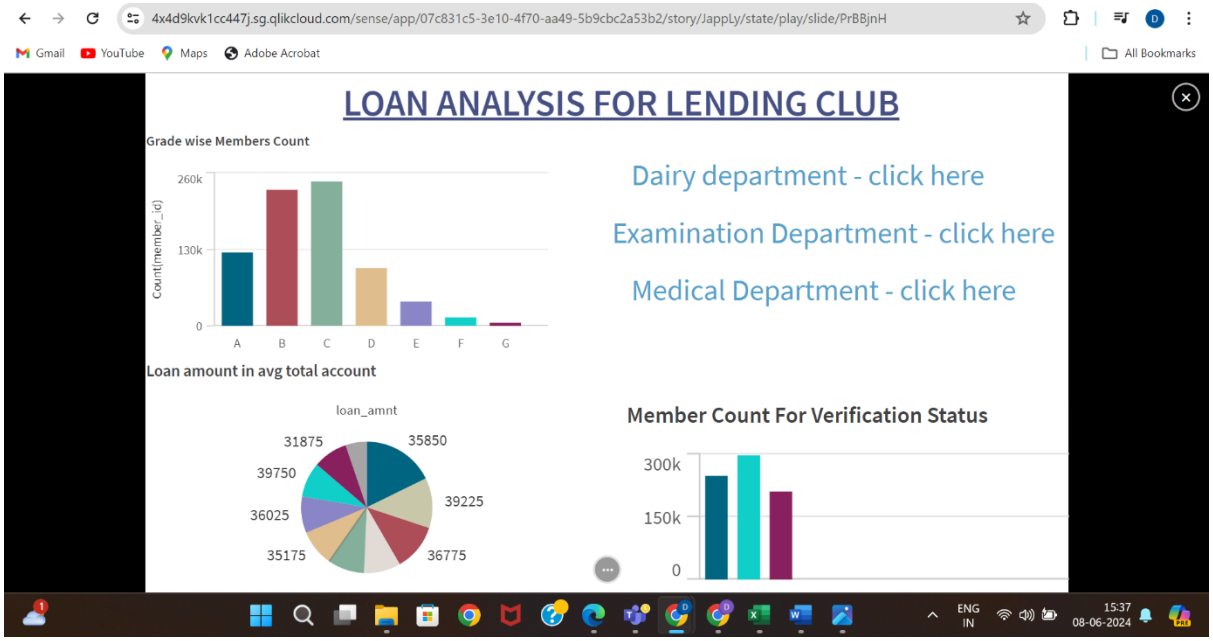
7.1 Story Creation

A data story is a way of presenting data and analysis in a narrative format, with the goal of making the information more engaging and easier to understand. A data story typically includes a clear introduction that sets the stage and explains the context for the data, a body that presents the data and analysis in a logical and systematic way, and a conclusion that summarizes the key findings and highlights their implications. Data stories can be told using a variety of mediums, such as reports, presentations, interactive visualizations, and videos.

Story 1



Story 2

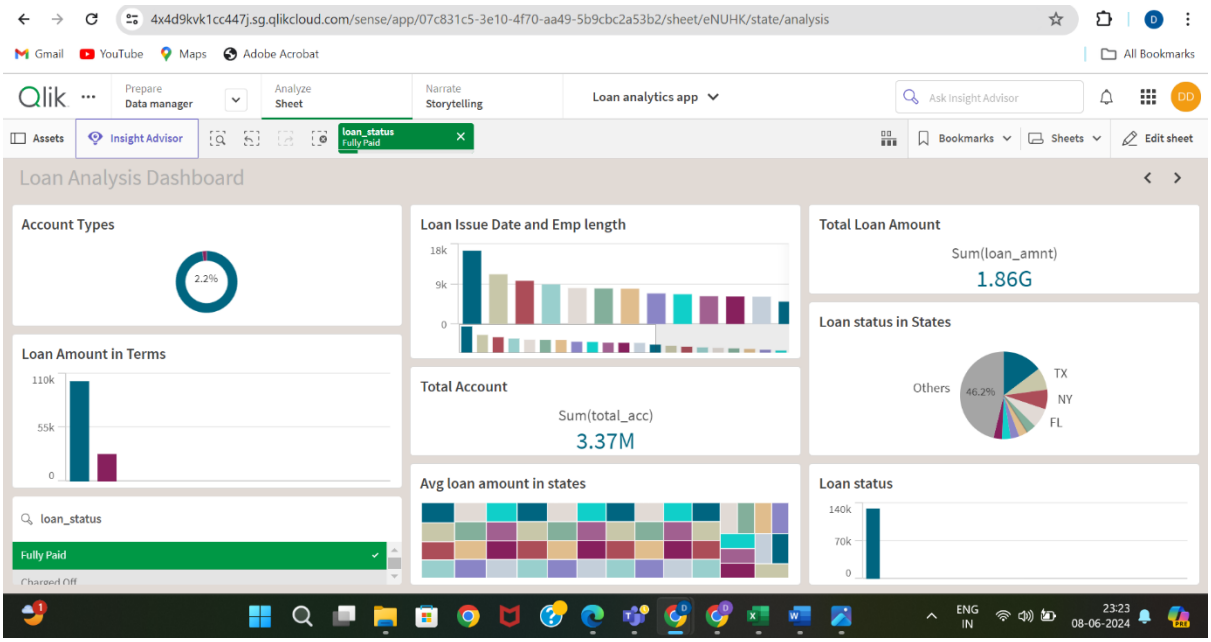


8 Performance Testing

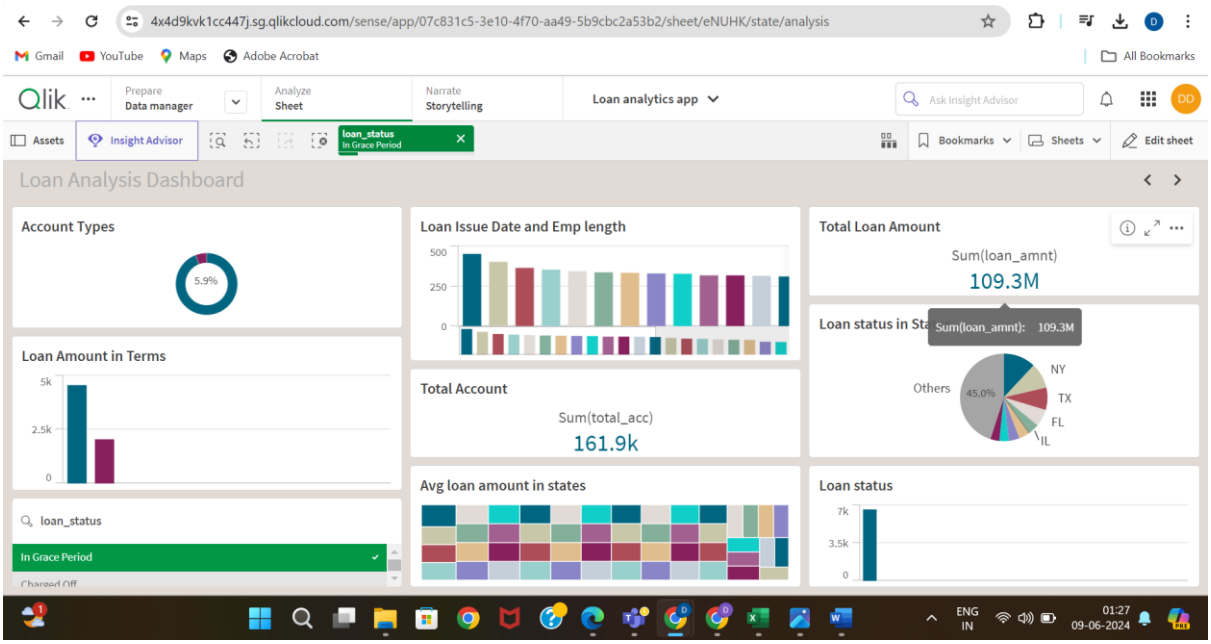
8.1 Amount of Data Rendered

8.2 Utilization of filters

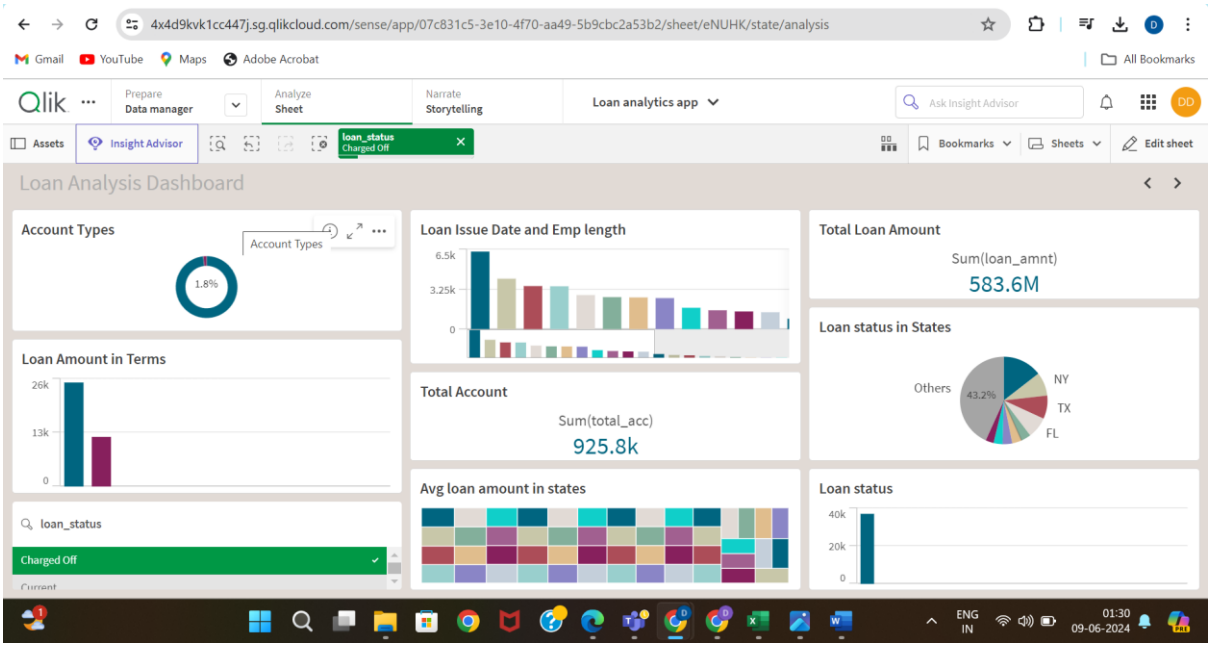
1) Filter is used to show data for loan status i.e fully paid , current , charged off , late etc.



The above filter shows for fully paid(Loan Status) one .

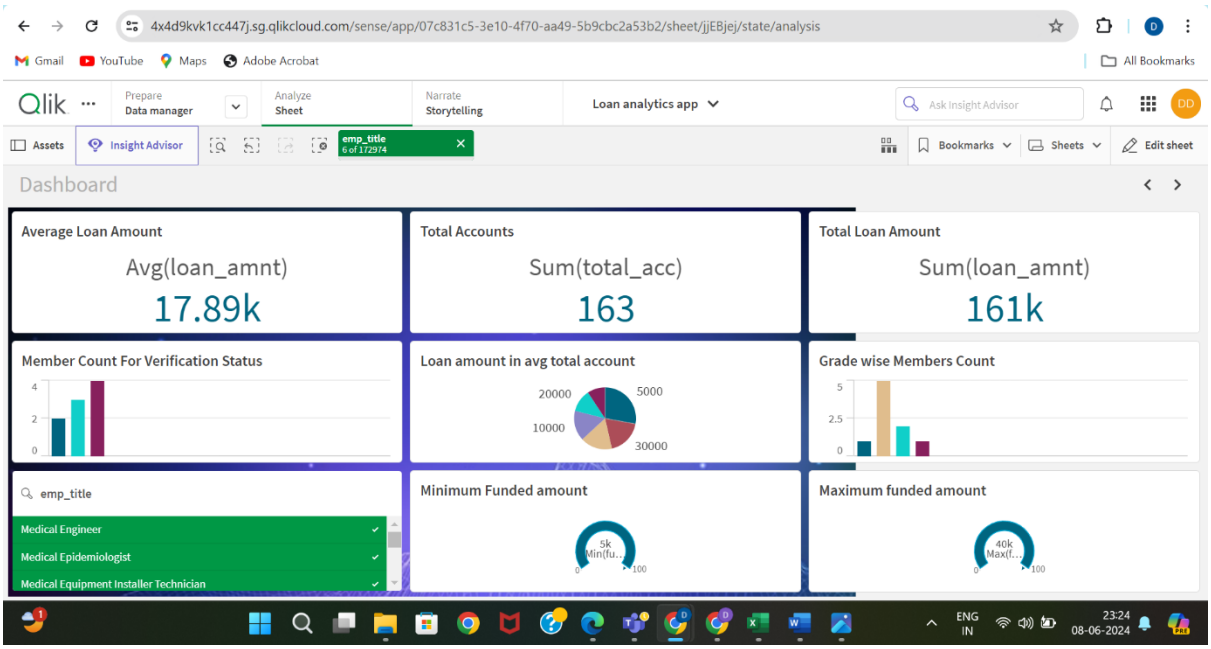


The above filter shows for In Grace period filter(Loan Status) one.

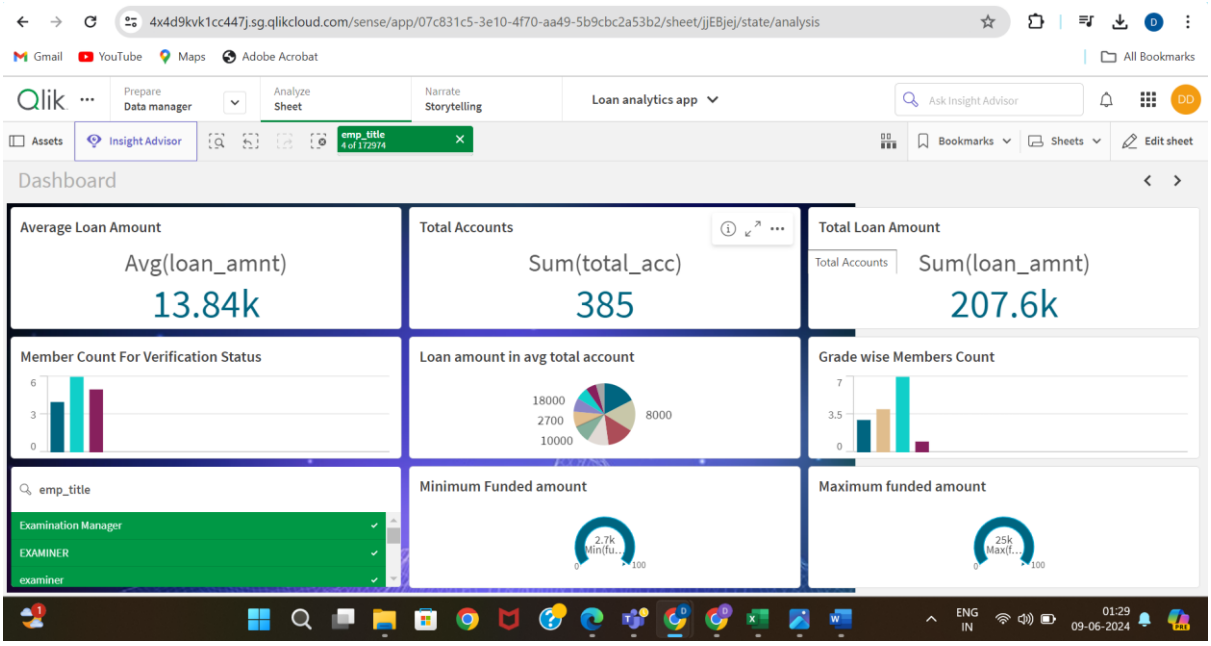


The above filter shows for Charged – Off period (Loan Status).

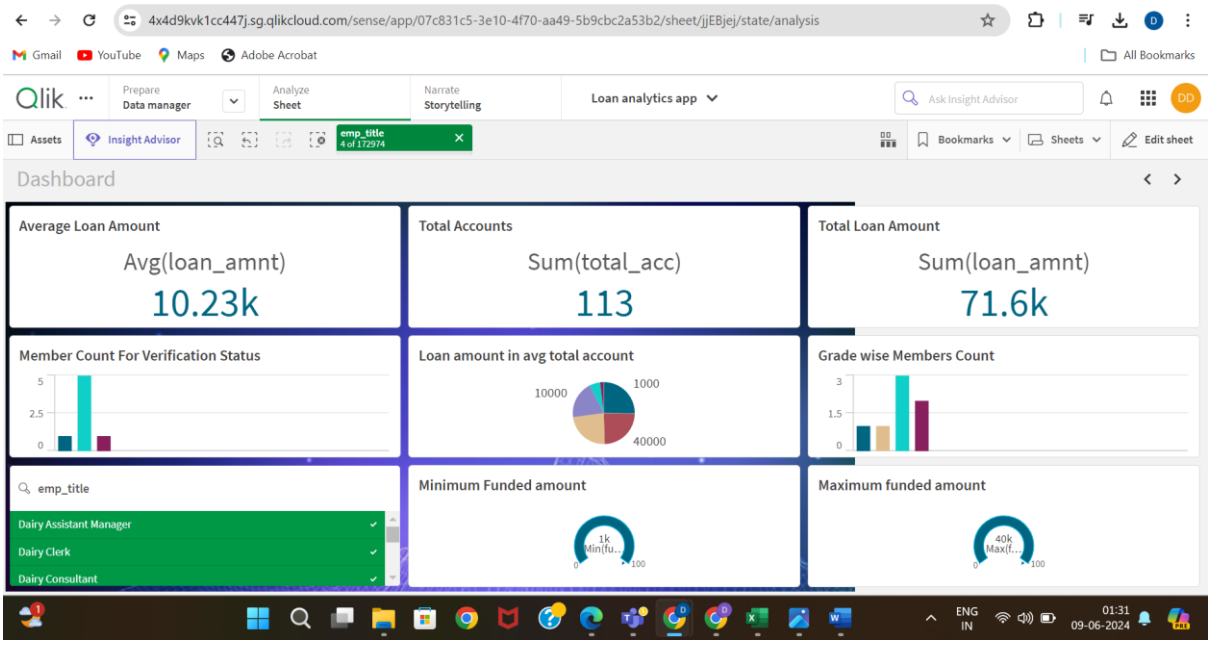
2) Filter is used to show data for employee title and accordingly the representation.



The above filter shows for Medical department (Employee Title).



The above filter shows for Examination department (Employee Title).



The above filter shows for Dairy department (Employee Title).